

Super-Resolution of Satellite Images: Feasibility of Deep Learning Techniques

Elena N. Akimova^{1, 2, a)}, Artem A. Deikov²

¹*N. N. Krasovskii Institute of Mathematics and Mechanics, Ural Branch of RAS, Ekaterinburg, Russia*

²*Yeltsin Ural Federal University, Yekaterinburg, Russia*

^{a)} Corresponding author: aen15@yandex.ru

Abstract. The work is devoted to studying the feasibility of applying the convolutional neural networks with deep learning to the problems of super-resolution of satellite images. The main aim is to enhance the image details and delete the artifacts. The algorithms for resolution enhancement were studied. The training set of satellite images was prepared. The neural network was constructed and trained using the PyTorch library for the Python language and the NVIDIA Tesla K40m graphics processors. Comparison of constructed network with the classic interpolation algorithms was carried out for the reference satellite images. It was shown that the neural network gives a better quality of the images.

Keywords: Super-resolution, PyTorch, CUDA, Python, neural network, deep learning.

INTRODUCTION

Remote sensing is the acquisition of information about an object or phenomenon without making physical contact with the object and, thus, in contrast to on-site observation, especially of the Earth. Remote sensing data is a reliable source of information for the global study of the Earth.

The most widely used civilian remote sensing tools are satellite imaging systems of the optical wavelength range. When processing data from such systems, the following problems often arise:

1. Visual enhancement;
2. Frame aggregation.

The paper proposes the development and analysis of applicability of the deep learning methods to achieve high resolution and improve the visual properties of images obtained by using the remote sensing.

High resolution means the greater pixel density and, hence, more precise detail. Different objects are easier to distinguish from each other by the satellite imagery if they have high resolution.

METHODS FOR INCREASING THE RESOLUTION

The single-frame spatial methods are divided into linear and nonlinear. The single-frame nonlinear spatial methods include: the gradient method; NEDI, and regularization method.

The single-frame spatial linear methods include [1–5] the nearest neighbor algorithm, bilinear interpolation, bicubic interpolation (Fig.1), gaussian interpolation, “Perfect” interpolation, and Lanczos method.

The interpolation method is a convolution

$$f(x, y) = \sum_{i,j=-\infty}^{+\infty} F(i, j)K(i - x)K(i - y), \quad (1)$$

where $F(i, j)$ is the pixel intensity in the image; K is the mask applied to the image.

For the nearest neighbors interpolation $K(t) = \begin{cases} 1 - |t|, & t \in [-1; 1), \\ 0, & t \notin [-1; 1), \end{cases}$ for the bilinear interpolation $K(t) = \begin{cases} (a + 2)|t|^3 - (a + 3)|t|^2 + 1, & |t| \leq 1, \\ a|t|^3 - 5a|t|^2 + 8a|t| - 4a, & 1 < |t| < 2, \\ 0, & |t| \geq 2. \end{cases}$ for the bicubic interpolation $K(t) = \begin{cases} 1, & t \in [-0.5; 0.5), \\ 0, & t \notin [-0.5; 0.5), \end{cases}$ $a = -0.5$.



FIGURE 1. Spatial linear methods (source, nearest neighbors algorithm, bilinear interpolation, and bicubic interpolation).

The use of these image scaling methods leads to appearance of artifacts. The most common artifacts are associated with the distortion of high-frequency information: the Gibbs phenomenon, Aliasing, blur effect. These artifacts are shown in Fig. 2.

Achieving a good result in suppressing the artifacts is possible by using the of post-processing, as well as by using the nonlinear processing methods; although the complete disappearance of artifacts cannot be achieved.

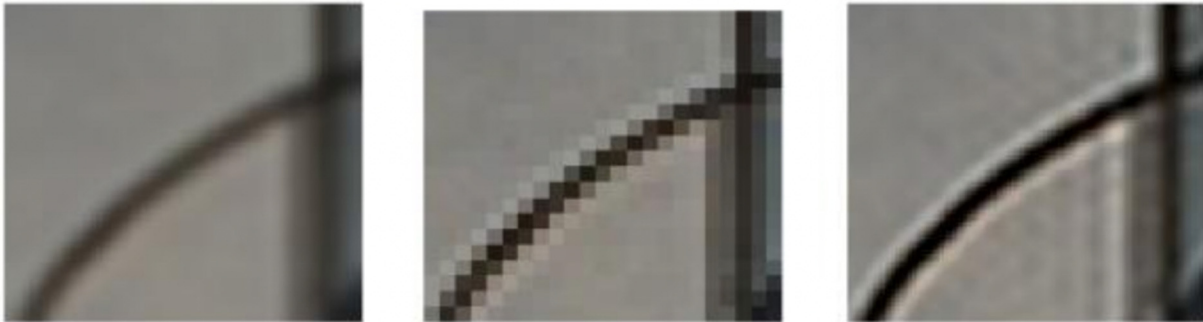


FIGURE 2. Artifacts caused by the image scaling (blur effect, Aliasing, and Gibbs phenomenon).

DEEP LEARNING AND NUMERICAL EXPERIMENTS

Deep learning is part of a broader family of machine learning methods based on the artificial neural networks. The learning can be supervised, semi-supervised, or unsupervised.

In the deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks most commonly applied to analyzing visual imagery.

The convolutional neural networks have been successfully used in image processing problems for the last century. But only recently, they have gained explosive popularity due to the significant superiority over other methods of image classification. A significant role in this progress was played by the effective implementation of training on modern powerful graphics processors and by easy access to an abundance of data for training models. It is not surprising that, in the problem of super-resolution of images, methods based on the CNN soon showed excellent results.

Our model of an artificial neural network (ANN) is based on the ResNet architecture described in [6] and is a multi-layer convolutional neural network that increases the resolution of the processed image for 4 times. The mean square error was used to estimate the accuracy.

Description of the model of an artificial neural network of with 13 layers.

- A 3-channel color image is fed to the input of the ANN.
- The layer 1 is a convolution layer with 3 input channels and 64 output channels.
- The layer 2 is an activation function.
- The layers from 3 to 12 are the core ones that have been introduced in paper [7].
- The layer 13 is a convolution one with 64 input channels and 3 output channels.

The neural network was implemented using the Python programming language and the PyTorch machine-learning library, which allows using the computing power of a video card for calculations. To work with the raster graphics, the Python Imaging Library (PIL) is used.

To create a training sample, the satellite images were taken the Landsat-4,5,7,8 and Sentinel-2 satellites were downloaded. Access is provided by the US Geological Survey through <https://glovis.usgs.gov/> and by the Copernicus Project.

From the obtained 20 satellite images of different locations, various objects such as road sections, artificial structures, airplanes, forests, and hills in the resolution of $N \times N$ pixels were separately identified. All the resulting images of objects were reduced to the size of 256x256 pixels. The database size was 2098 images of various objects.

The neural network was trained using the collected data set, which were divided into the training sample of 1888 images and the test sample of 210 images. The training took place on the Uran supercomputer using the 4 GPU of the NVIDIA Tesla K40m and started with a 4-fold reduction in images and 32 random images on the network input in batches. The training time of one epoch was about 107 seconds.

We also conducted tests on training on the CPU; on average, the training time was 3972 seconds.

The total network has studied 100 epochs. The intermediate learning outcomes are shown in Fig. **Error! Reference source not found.3.**



FIGURE 3. Result of the network, after 1 and 7 epochs of learning.

The peak signal-to-noise ratio (often abbreviated PSNR) is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, the PSNR is most easily defined via the mean square error (MSE). The MSE is defined as

$$MSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y'_i - y_i)^2}, \quad (2)$$

where N is the number of pixels in the image; y'_i is the value of the i -th pixel of the converted image; y_i is the value of the i -th pixel of the original image.

The PSNR is defined as

$$SNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right), \quad (3)$$

where MAX_I is the largest value that is taken by one pixel.

To assess the results, the satellite images of Ekaterinburg dated November 15, 2015, and Kazakhstan ones dated August 12, 2011 were used provided by the Center for Space Monitoring of the Ural Federal University.

These images were taken by the following spacecraft:

- SPOT-4 [8] (1998-2013) about 450 thousand pictures per year with the pixel size 10-20 m.
- DMC-2G [8] (since 2009) and till nowadays about 3 thousand pictures per year with the pixel size 22 m.

The results obtained in the course of the algorithm application are summarized in Table 1. The neural network showed a fairly good result, and with an increase in the time and epoch of learning, one can achieve results higher than those shown in the table.

TABLE 1. Average value PSNR

Metric	Bicubic interpolation	Neural network
PSNR, dB	28.9	29.7

The results of the algorithm application are presented in Fig. 4. In these figures, we can observe not only improvements in resolution but, also, the color correction.



FIGURE 4. The result of the network using the left image shows the result of the work of the bicubic interpolation, the right one shows the result of the neural network application.

CONCLUSION

In this work, the following results were obtained:

- Application of the convolutional neural networks with deep learning to increase the resolution of satellite images leads to increased image detail without the appearance of artifacts.
- An educational sample of space images was prepared.
- Using the PyTorch library, the Python language, and the Tesla K40m graphics processors, the neural network has been built and trained.
- Comparison of the constructed convolutional neural networks with the classical interpolation methods on reference satellite images was carried out. The improvement of the images quality was obtained by the constructed neural network.

REFERENCES

1. K. S. Markelov. Super-resolution digital image enhancement model. Moscow Engineering Bulletin. 2013. Vol. 3. P. 525-542.
2. E. I. Petrovichev. Computer Graphics: Tutorial. Moscow. MGGU. 2003.
3. Gonzalez R., Woods R. Digital image processing. Moscow. Technosphere. 2005.
4. "Perfect" interpolation — <https://siblec.ru/telekommunikatsii/radiosistemy-peredachi-informatsii#3.3>
5. A.V. Nasonov. Regularizing methods for increasing resolution and super resolution. Moscow. 2011.
6. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR 2016.
7. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network.
8. V. G. Kobernichenko et al. Earth Remote Sensing Data Processing: Practical Aspects: A Training Manual. Ekaterinburg. 2013.